Weekly Report

Lu Junhua

2015 年 8 月 9 日

This week, a preliminary result of statical data in Gongan. Although we did this job before, the data we used that time is not accurate and the volume of data was small. This time almost one hundred thousand persons were included. The crime rate is around 1.5%. Here, we consider anyone who didn't have crime records as innocent(although he may leave a mark in 打防控people database.)

	crime		total
crime_hat	0	1	
0	91814	973	92787
	93.89%	65.84%	93.46%
1	5979	513	6492
	6.11%	34.52%	6.54%
total	97793	1486	99279
	100.00	100.00	100.00

crime is the precise value and crime_hat is the estimated value. This table is the results on Thursday, on Friday we get a better result. From the table we can see that about one third of the crimes are correctly pick out, and on Friday, after some revisions on the algorithms and codes, we got a better result.

Here, categorical data has been processed as dummy variables, thanks to STATA which is recommended by Prof. He. And temporarily the time-variant variable are not used, we may have a test on this next week, from the basic regression method combined with time axis.

I put all the data from csv to oracle database. I came across several problems and most of them has been solved. One remain is that there too many \t \n in the content of a crime records which contributes to the error in positions of attribute. I had an idea for solving this is that: first remove all the quotes in the text and then, add two quotes on the beginning and end of the text.

For more details, please refer to the Chinese edition of weekly report on Gongan Data. And the picture on page 2 reveal another result of the convincing of variables, smaller the p, better reliability.

On coding pactice, I read the *Hierarchical Edge Bundles: Visualization of Adjacency Relations in Hierarchical Data* by Holten. It's an elegant visual expression of hierarchical data. It uses B-spline curves and different hierarchical nodes.

Besides, I have learned more about D3.js. I have known a new method named path. With svg path, we can draw smooth curves which comes through all the given points. Bezier, b-spline, and other specified method is ok. And also, using D3's enter() and function() functions, we can dynamically draw these

curves, which I think is a solution to stacked graph. However I haven’t understood the fill function of it and I may had a implementation next week.

		Robust				
crime	Coef.	Std.Err.	z	P>|z|	[95% Conf.	Interval]
cage(年龄)	-0.1133517	0.0041283	-27.46	0.000	-0.121443	-0.1052604
cage2(年龄²)	0.0000762	0.00000206	36.96	0.000	0.0000721	0.0000802
_Imarrital_2(初婚 复婚)	1.1344	0.1176491	9.64	0.000	0.9038119	1.364988
_Imarrital_4(离婚)	2.609452	0.2390792	10.91	0.000	2.140865	3.078038
_Imarrital_5(丧偶)	2.86031	0.04389773	6.52	0.000	1.99993	3.72069
_Imarrital_6(未婚)	-0.2853385	0.1120975	-2.55	0.011	-0.5050455	-0.06563115
_Iedu_100	1.576301	0.1457449	10.82	0.000	1.290646	1.861955
_Iedu_200	1.404292	0.1454378	9.66	0.000	1.119239	1.689345
_Iedu_300	0.9730494	0.1533659	6.34	0.000	0.6724577	1.273641
_Iedu_400	-0.310695	0.2637498	-1.18	0.239	-0.8276352	0.2062452
_Iedu_500	-0.5934555	0.2504116	-2.37	0.018	-1.084253	-0.1026578
temp(暂住)	-0.9076115	0.1027456	-8.83	0.000	-1.108989	-0.7062338
_Iorigin_2(省内城区)	0.59135	0.5703547	1.04	0.300	-0.5265246	1.709225
_Iorigin_3(省内市辖镇)	0.4055489	0.1210854	3.35	0.001	0.1682259	0.6428719
_Iorigin_4(省内县辖镇)	0.2311314	0.1349964	1.71	0.054	-0.0334568	0.4957196
_Iorigin_5(省内乡村)	0.8536197	0.1908333	4.47	0.000	0.4795934	1.227646
_Iorigin_6(省外城区)	1.013137	0.4028077	2.52	0.012	0.2236479	1.802625
_Iorigin_7(省外市辖镇)	-0.451966	0.1396474	-3.27	0.001	-0.7309005	-0.1834927
_Iorigin_8(省外县辖镇)	-0.2899849	0.1295688	-2.24	0.025	-0.5439351	-0.0360347
_Iorigin_9(省外乡村)	0.8027473	0.1504899	5.33	0.000	0.5077925	1.097702
female	-1.53735	0.0700647	-21.94	0.000	-1.674674	-1.400026
_cons	-0.1122316	0.2030404	-0.55	0.580	-0.5101834	0.2857202